A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation

A Probabilistic Reconciliation of Coherence-Driven and Centering-Driven Theories of Pronoun Interpretation

Andrew Kehler
University of California, San Diego

Hannah Rohde
University of Edinburgh

First Version Submitted 2012-05-25; Final Version Submitted 2012-02-24

Abstract Two classic theories of pronoun interpretation have each sought to specify the relationship between pronoun use and discourse coherence, but make seemingly irreconcilable claims. According to Hobbs (1979, 1990), pronoun interpretation is not governed by an independent mechanism, but instead comes about as a by-product of utilizing world knowledge during the inferential establishment of discourse coherence relations. Factors pertaining to the grammatical form and information structure of utterances do not come into play. According to Centering Theory (Grosz, Joshi & Weinstein 1995; inter alia), on the other hand, pronoun interpretation is predominantly determined by information structural relationships within and between utterances (e.g., topic transitions) and the grammatical roles occupied by potential referents. Factors pertaining to world knowledge and the establishment of informational coherence relations do not come into play.

In this paper, we describe a series of psycholinguistic experiments that ultimately suggest a reconciliation of these diverse approaches. These experiments reveal a definitive role for coherence relationships of the Hobbsian sort, demonstrating that pronoun interpretation is affected by (i) probabilistic expectations that comprehenders have about what coherence relationships will ensue, and (ii) their expectations about what entities will be mentioned next which, crucially, are conditioned on those coherence relationships. However, these experiments also reveal a role played by the topichood status of potential referents. These data are reconciled by a probabilistic model that combines the comprehender's prior coherence-driven expectations about what entities will be referred to next and Centering-driven likelihoods that govern the speaker's choice of referential form. The approach therefore situates pronoun interpretation within a larger body of work in psycholinguistics, according to which language interpretation results when top-down predictions about the ensuing message meet bottom-up linguistic evidence.

Keywords: pronoun interpretation, discourse coherence

1 Introduction

Our goal in this paper is to examine the mechanisms that underlie discourse processing, viewed through the eyes of pronoun interpretation. So let us start with a brief characterization of the pronoun interpretation problem, and of some of the early proposals put forth for dealing with it. Consider the following passages:

(1) a. Mitt narrowly defeated Rick, and the press promptly followed him to the next primary state. [him = Mitt]

b. Rick was narrowly defeated by Mitt, and the press promptly followed him to the next primary state. [him = Rick]

c. Mitt narrowly defeated Rick, and Newt absolutely trounced him. [him = Rick]

d. Mitt narrowly defeated Rick, and he quickly demanded a recount. [he = Rick]

If you poll native speakers for judgments about the reference of the pronoun in (1a), most will tell you that *him* refers to Mitt. However, if you ask instead about passage (1b), those same speakers will tell you the pronoun refers to Rick. Such facts have in part caused researchers to posit a SUBJECT ASSIGNMENT STRATEGY in pronoun interpretation (Frederiksen 1981; Crawley, Stevenson & Kleinman 1990; inter alia), whereby comprehenders preferentially assign a pronoun to the entity mentioned in the subject position of the previous clause.[1] After all, the contexts leading up to the pronoun in passages (1a-b) are semantically invariant; all that changes between (1a) and (1b) is the voice used in the first clause. The switch from active to passive voice changed the entity that appears in subject position, and the pronoun interpretation preference went with it. So what else could it be, if not a structural preference tied to subject position?

This reasoning is well and good until we encounter examples like (1c). Most informants will now tell you that the pronoun *him* refers to Rick. So much for the subject assignment strategy that we thought (1a-b) established. Such data has caused researchers to posit a GRAMMATICAL ROLE PARALLELISM bias (Smyth 1994; Chambers & Smyth 1998; inter alia), whereby a pronoun prefers a referent that occupies the same grammatical role in another clause. While intuitively appealing

1 Other researchers (Gernsbacher & Hargreaves 1988; Gernsbacher, Hargreaves & Beeman 1989) have instead taken such data to be evidence for a FIRST-MENTION ADVANTAGE. See Järvikivi, van Gompel, Hyönä & Bertram (2005) for a comparison of the two proposals.

in this example, one nonetheless wonders why the same preference did not similarly apply in passage (1a), in which a pronoun in object position has an available antecedent in object position.

Muddying the waters further are examples like (1d). Here, informants will favor Rick as the referent of *he*, even though the subject assignment strategy and parallel grammatical role bias both agree on Mitt. Such examples have caused researchers to claim that pronoun interpretation is determined primarily by world knowledge and reasoning (Hobbs 1979). This might seem to be an obvious explanation: After all, who's more likely to demand a recount, the winner or loser of an election? But then one wonders why a purely syntactic manipulation as seen in (1a-b) – where again, the examples are invariant with respect to semantics and world knowledge – could change the preferred referent of a pronoun.

When we consider their highly similar morphosyntactic properties, it is not immediately obvious why examples (1a-d) appear to favor different strategies for interpreting pronouns. If we take a closer look at them, however, we see that they are not entirely alike in a deeper respect: specifically in terms of what we might informally refer to as the semantic 'glue' that binds the clauses in each case together to form a coherent discourse. For example, passages (1a-b) seem to tell a story as a narration, i.e. a series of contingent, temporally-connected events. Passage (1c), on the other hand, coheres for reasons that appear to be based on semantic parallelism: Two similar events happened, and unlike (1a-b), we cannot even be sure as to the order in which the events occurred. Passage (1d) is likewise different from both of these, being bound by causality, with the first clause describing a cause and the second its effect. It therefore stands to reason that perhaps different semantic relationships can affect pronoun interpretation in different ways (cf. Hobbs 1979, 1990).

While this may sound plausible, it is not the direction that pronoun research has historically taken. Instead, work in both psycholinguistics and computational linguistics has often assumed what we will (uncharitably) refer to as the SMASH algorithm as a model of interpretation (Kehler 2007):

Search: Collect possible referents (within some contextual window)

Match: Filter out those referents that fail 'hard' morphosyntactic constraints (number, gender, person, binding)

And Select using Heuristics: Select a referent based on some combination of 'soft' constraints (grammatical role, grammatical parallelism, thematic role, referential form, ...)

Under this model, encountering a pronoun triggers a fairly complex process in the mind of the comprehender. First, a search is initiated to collect possible referents.

Then a series of constraints is applied to eliminate potential referents, based on number, gender, binding theory, and so forth. Then, with what's left, a computation ensues in order to determine which of the potential referents is the most likely one. In heuristic theories, this is just the referent that the heuristic singles out (e.g., the subject assignment strategy picks the referent that was mentioned in the subject position of the previous sentence, if available). In other models, the underlying computation is more complex. For instance, in the Centering-based interpretation model of Brennan, Friedman & Pollard (1987) and Walker, Iida & Cote (1994), all possible combinations of referent assignments that are consistent with Centering rules are considered and ranked according to four ordered transition types, which in turn have to be computed from the configurational properties of the forward and backward looking centers. Other multi-factorial theories (Stevenson, Nelson & Stenning 1995; Arnold 2001) employ a competition model style framework in which weighted sets of varied sorts of featural information (grammatical roles, thematic roles, and so forth) are used to determine the most likely referent.

From this picture emerges what one might call the BIG QUESTION: If the interpretation of pronouns is really this involved, why would any speaker do the disservice to her addressee of using one? After all, in doing so she is choosing to use an ambiguous expression in lieu of an unambiguous one, such as a proper name. Yet we have reason to believe, both intuitively and experimentally, that the felicitous use of a pronoun does not particularly hinder discourse comprehension, and in some cases even facilitates it (Gordon, Grosz & Gilliom 1993). So what we want from a theory is not merely a list of so-called 'preferences' or 'heuristics' that people purportedly employ when interpreting pronouns; after all, without any underlying explanation offered, these preferences do not do much more than restate statistical generalizations found in the data. What we want instead is a deeper explanation for why we find such patterns, and indeed why different preferences seem to prevail in different contextual circumstances (as seen in 1a-d).

We believe that the complexity we find in pronoun interpretation data, therefore, should not be met with an equally complicated theory of it. Instead, we suspect that the complexity of the data arises not from the process of pronoun interpretation itself, but instead from its coming into contact with the operation of discourse-level inferential processes that we already independently know to exist: that is, that complexity results of the integration of simpler parts. We will thus carry forth with the hypothesis that pronoun interpretation is simpler than commonly assumed, and ask: *What would the discourse processing architecture have to look like to allow for a fairly simple theory of pronoun interpretation?* In this way, we study pronouns not merely to better understand them as linguistic objects, but to use them the same way that biologists use fruit flies: In our case, to provide a window into larger questions about the architecture and workings of the discourse processing mechanism.

2 Two Theories of Discourse Coherence

In the remainder of the paper, we will focus on two analyses that have attempted to explain pronoun behavior in terms of a larger theory of discourse interpretation and coherence; approaches that make opposite and seemingly irreconcilable claims. The first is the Centering Theory of Grosz et al. (1995). Centering is a fairly complex theory that comes in a variety of instantiations, and hence a full description would take us somewhat afield. We will instead focus on what we take to be the primary claim of the theory, captured in the following quotes from the preface of Grosz et al.:

> Certain entities in an utterance are more central than others and this property imposes constraints on a speaker's use of different types of referring expressions... The coherence of a discourse is affected by the compatibility between centering properties of an utterance and choice of referring expression. (p. 203)

Note that this is a *production-oriented* claim, in that it posits the existence of constraints on speakers in their choice of referring expressions that impact the coherence of the discourse in which they appear. Specifically regarding a speaker's choice to employ a pronoun, they further say:

> ...the use of a pronoun to realize the C_b signals the hearer that the speaker is continuing to talk about the same thing. (p. 214)

Centering implements these ideas through the use of several constructs and rules. The constructs include a partially-ordered list of forward-looking centers (C_f) which, roughly speaking, contains all of the entities referred to in the utterance. Among these is the (single) backward-looking center (C_b), which by definition is the element in the current utterance that is most highly ranked on the C_f of the previous utterance. The C_b can be thought of as the topic of the utterance; in line with the aforementioned quote, a rule of Centering states that if any element in the current utterance is pronominalized, the topic should be. There are also rules that rank types of interclausal transitions that are defined by the C_b and C_f which have been used in Centering-based pronoun interpretation algorithms (Brennan et al. 1987; Walker et al. 1994), which will not be of use to us here (see Section 4).

While Centering in principle allows for a variety of factors to influence the ranking of the C_f, one that is most commonly appealed to is grammatical role, assuming the ordering provided by the obliqueness hierarchy *Subject* > *Object(s)* > *Other* (Grosz et al. 1995; Brennan et al. 1987; Walker et al. 1994; inter alia).[2] With

2 This statement applies to work on English. For discussions of Cf ranking in other languages, see e.g. Kameyama (1986) and Walker et al. (1994) for Japanese, Turan (1998) and Hoffman (1998) for

this in mind, one can see how a suitably-constructed Centering-based algorithm could conceivably explain why a difference in choice of syntactic form such as (1a-b), repeated below in (2), could ultimately affect how the ensuing pronoun is interpreted, in so far as different entities are mentioned from subject position.

(2) a. Mitt narrowly defeated Rick, and the press promptly followed him to the next primary state. [him = Mitt]

b. Rick was narrowly defeated by Mitt, and the press promptly followed him to the next primary state. [him = Rick]

Importantly, the relevant constructs of Centering are defined in terms of morpho-grammatical properties of the discourse. Semantics, world knowledge, and inference do not come into play.

A second and diametrically-opposed approach is the coherence-driven theory of Hobbs (1979). In this approach, establishing discourse coherence is achieved by making the inferences necessary to meet the constraints imposed by one of a set of COHERENCE RELATIONS. Consider a passage like (3):

(3) Mitt flew to San Diego this weekend. He has family there.

A typical comprehender will not interpret the two statements in (3) as independent facts about Mitt, but instead infer that the second sentence expresses the reason for the eventuality described in the first. Although this relation is nowhere expressed in the passage, comprehenders will typically make inferences necessary to establish a way in which the statements are relevant to one another. In the case of example (3), this relationship is characterized as an EXPLANATION relation, informally defined as follows:[3]

Explanation: Infer that the second sentence describes a cause or reason for the eventuality described in the first sentence.

Explanation is only one such relation; several other common ones include:

Occasion: Infer a change of state from the second sentence, taking its initial state to be the final state of the eventuality described in the first sentence.

Mitt flew to San Diego this weekend. He took a taxi from the airport to his house.

Turkish, Di Eugenio (1998) for Italian, Rambow (1993) and Strube & Hahn (1999) for German, and Hedberg (2010) for Kaqchikel Mayan.

3 See Kehler (2002) for more formal definitions of this relation as well as the others discussed below.

Elaboration: Infer that both sentences provide descriptions of the same eventuality.

Mitt flew to San Diego this weekend. He took a private jet into Lindbergh field.

Result: Infer that the first sentence describes a cause or reason for the eventuality described in the second sentence.

Mitt flew to San Diego this weekend. He was therefore able to visit several high-profile campaign donors.

Violated Expectation: Infer that the second sentence describes an unexpected result of the eventuality described in the first sentence.

Mitt flew to San Diego this weekend. He wasn't able to visit any high-profile campaign donors, however.

Parallel: Infer that the first and second sentences express similar eventualities, as if each provides a partial answer to a common question.

Mitt flew to San Diego this weekend. Rick stayed in Kansas to campaign.

In Hobbs's analysis, pronoun interpretation is not an independent process at all, but is instead a by-product of more general reasoning about the most likely meaning of a discourse, including the establishment of discourse coherence. Pronouns are modeled as free variables which become bound during these inference processes; potential referents of pronouns are therefore those which result in valid proofs of discourse coherence. The analysis therefore offers a principled account of why examples like (4) with follow-ons (a) and (b), adapted from Winograd (1972), lead to two different interpretations for the pronoun *they*.

(4) The city council denied the demonstrators a permit because

a. they *feared* violence.

b. they *advocated* violence.

The connective *because* in each case explicitly indicates an Explanation relation. As such, the pronoun receives the assignment necessary to establish this relation, in light of the fact that the city council fearing violence is a more likely cause than the demonstrators fearing violence (4a), but that the demonstrators advocating violence is a more likely cause than the city council advocating violence (4b). Notably, these different interpretations result even though the two sentence pairs have identical syntactic configurations; indeed, the only difference between them is the verb that comes after the pronoun in the second clause. The Centering analysis offers no explanation for this difference (also recall example 1d). Indeed, the Hobbs

approach is in many ways the opposite of Centering: Semantics, world knowledge, and inference drive interpretation, whereas choice of linguistic form (syntactic, referential) does not come into play.

The rest of the paper is organized as follows. We first describe the results of a series of experiments that demonstrate that pronoun interpretation is highly sensitive to coherence relations of the Hobbsian sort, and which find no obvious explanation in a Centering analysis. We then follow by presenting other evidence, however, that remains unexplained by the Hobbsian story, and which in fact suggests a role for a Centering-based constraint. We culminate the discussion by presenting a simple probabilistic model that integrates aspects of both approaches, and describe an experiment that examines predictions of the model. The end result is an analysis that, while reconciling the core tenets of both theories, is also argued to improve upon these theories in a number of important respects.

3 Experiments

Our examination of these approaches to coherence and their relationship to pronoun interpretation starts with a set of intriguing data from a series of passage completion experiments conducted by Stevenson, Crawley & Kleinman (1994). Of the various different context types they considered, we will focus on those that utilized transfer-of-possession verbs, as in (5–6):

(5) John seized the comic from Bill. He _____

(6) John passed the comic to Bill. He _____

Both (5) and (6) describe a transfer of possession, but they differ with respect to where the fillers of the Source and Goal thematic roles occur: In (5) the Goal is the subject, whereas in (6) it appears as the object of the prepositional phrase. Stevenson et al. instructed experimental participants to complete passages like (5–6) and then coded the intended interpretations of the pronouns. For Goal-Source contexts like (5), they found an overwhelming bias (84.6%) to interpret the pronoun to refer to the subject/Goal (i.e., *John*). For contexts like (6), however, the distribution was almost even: 51% to the subject/Source and 49% to the non-subject/Goal. This latter result is somewhat surprising, since an object-of-PP referent at the end of the sentence – normally a relatively dispreferred position for a pronominal referent to appear in – would not typically be expected to compete with the sentential subject, especially in light of the previously discussed subject assignment and parallel grammatical role biases, both of which favor the Source.

Based on results such as these, Stevenson et al. posited that there are two biases that come into play for pronoun interpretation. The first is a grammatical

role preference that ranks subjects over non-subjects, as previous researchers have proposed. The second is a thematic role bias that similarly ranks occupants of the Goal role over occupants of the Source. These two preferences agree for (5), where the subject is also the Goal, and hence one finds a large proportion of assignments to the subject. The preferences conflict in (6), however – John is the subject but Bill is the Goal – hence why one finds an even distribution of referents.

Stevenson et al. offered an explanation for the Goal bias that went beyond it merely being a heuristic preference. They hypothesized that after interpreting an utterance, comprehenders focus their attention on the end state of the described eventuality in their mental model of the discourse as they proceed to interpret the subsequent utterance. Whereas for many event types one would expect the initiator (e.g., the Agent) to remain central throughout the event structure, it stands to reason that there would be a shift for transfer-of-possession events. That is, while we might expect the Source to be central to the initial state of the event, intuitively this centrality shifts to the Goal – i.e., the recipient of the object of transfer – at the point at which the end state is reached. In Stevenson's model, this end-state focus is a "top-down" bias that yields predictions about what event participants are likely to be mentioned next (henceforth referred to as the NEXT-MENTION bias), regardless of the form of reference used. The aforementioned subject bias, on the other hand, results from a "bottom-up" strategy that is triggered specifically by the occurrence of a pronoun.

Stevenson et al.'s theory therefore went beyond previous accounts in moving away from an approach based only on pronoun-specific heuristics. As we noted in Rohde, Kehler & Elman (2006) (see also Arnold (2001); Kehler, Kertz, Rohde & Elman (2008)), however, the proposed focus on the end state of the previous eventuality only makes sense for certain coherence relations with which the discourse could be felicitously continued. On the one hand, it makes perfect sense for Occasion relations, the definition of which is repeated below:

Occasion: Infer a change of state from the second sentence, taking its initial state to be the final state of the eventuality described in the first sentence.

Note that the inference procedure underlying the establishment of Occasion directly incorporates a focus on the end state of the previous eventuality: It is this state of affairs that is inferred to be the initial state for interpreting the next eventuality. In this sense, a focus on the previous end state is a natural consequence of establishing Occasion, without any further stipulation. One might similarly anticipate an end-state bias for Result relations, since effects are associated with the consequences of an event. On the other hand, similar logic fails to hold for many of the other relations. For instance, we would expect Explanation relations to incorporate a focus on the

initial state of the previous eventuality, as causes precede effects. Similar reasoning holds for Elaboration relations, which generally take the form of a redescription of the previous eventuality or aspects of it, and hence offer no reason to expect an end-state focus.

We therefore designed an experiment with two goals in mind. We first wanted to test the event-structure explanation for the Goal bias as compared to a simpler heuristic that uniformly ranks Goals over Sources. Second, assuming that the event structure proposal wins, we wanted to evaluate the predictions of the coherence account – in particular, that the Goal bias would be limited primarily to those coherence relations for which we would naturally expect to find it for independent reasons, such as Occasion and Result.

The experiment paired passages like (7), in which the verb occurs in the perfective form, with versions like (8), in which the imperfective form was used.

(7) John passed a comic to Bill. He _____

(8) John was passing a comic to Bill. He _____

Crucially, the thematic role bias predicts no effect of aspectual form, since the thematic roles are identical between examples (7) and (8). Thus, equivalent distributions of reference between (7) and (8) would support the thematic role bias. However, these examples are not equivalent with respect to event structure: By its nature, the imperfective creates a focus on the ongoing development of an event, and hence is incompatible with a focus on the end state. Because the Source is presumably still central to the event as it is ongoing, the event structure hypothesis predicts that we will find a greater percentage of Source interpretations in the imperfective condition than in the perfective condition.

The results supported the event structure hypothesis. As seen in Table 1, imperfective context sentences yielded significantly more Source interpretations than perfective sentences.[4]

	Perfective	Imperfective
Source Interpretation	.57	.80

Table 1 Proportion of Source Interpretations (Aspect Manipulation)

4 This paper synthesizes data presented in a number of other papers by the authors. Whereas we will forgo providing statistical significance results here, all comparisons reported on have been shown to be statistically reliable through application of appropriate tests. The results of these tests, discussions of methods and coding, and more detailed analyses of the data can be found in the referenced works.

The next question is whether the coherence relation that relates the continuation to the context sentence has any impact on the distribution of pronoun assignments. (So as to not overly complicate the discussion, we will restrict analysis to the perfective condition here and in future comparisons.) As can be seen in Table 2, coherence relations had a strong effect. As predicted, Occasion – the most common

	Frequency of Occurrence	Bias to Source
Occasion	.38	.18
Elaboration	.28	.98
Explanation	.18	.80
Violated Expectation	.08	.76
Result	.06	.08

Table 2 Distribution of Five Most Common Coherence Relations and their Pronoun Interpretation Biases (Perfective Condition)

relation in the data – exhibited a strong bias to the Goal. However the next two most common relations, Elaboration and Explanation, displayed a strong *Source* bias.[5] Thus, we see that although the overall pronoun bias is close to 50/50 (in our case, approximately a 57/43 split), there is nothing 50/50 about it when one conditions on coherence relations. The overall 50/50 profile masks the much stronger biases associated with the most common coherence relations.

Note that the data in Table 2 actually reflects two biases that are relevant to pronoun interpretation. The one represented in the right-hand column is the bias we just spoke of: The probability of a pronoun assignment conditioned on the coherence relation. The other bias, shown in the middle column, represents the fact that not all coherence relations are equiprobable. These two biases are associated in Equation (9), in which *referent* stands for an entity mentioned in a particular grammatical or thematic position, and *CR* stands for coherence relation):

$$(9)\quad P(pronoun = referent) = \sum_{CR \in CRs} P(CR) * P(pronoun = referent \mid CR)$$

That is, to compute the probability that a pronoun will corefer with the subject of the previous sentence, one can simply sum, over all coherence relations, the probability

5 This analysis is similar to one carried out by Arnold (2001), who coded a coarser three-way split between *cause*, *endpoint*, and *other* relations in data collected from a completion experiment using passages that lacked pronoun prompts. She likewise found differences in referential biases across continuation type.

of seeing that coherence relation multiplied by the probability of a subject reference given that coherence relation.

Now in and of itself, Equation (9) is not particularly interesting: It expresses a law of conditional probability. However, by conditioning on coherence relations, we are hypothesizing that, as suggested by the results of the previous experiment, the conditional probabilities $P(\textit{pronoun} = \textit{referent} \mid CR)$ vary across coherence relations. This, in turn, leads to a prediction: If some factor were to cause a change in the distribution of coherence relations $P(CR)$ without impacting the conditional probabilities $P(\textit{pronoun} = \textit{referent} \mid CR)$, then we should still see a change in the overall distribution of pronouns, despite the fact that $P(CR)$ does not mention reference at all. For instance, a shift towards a greater percentage of Source-biased relations (e.g., Explanation, Elaboration) in the data should lead to a greater percentage of pronominal references to the Source.

As an example, this account makes a straightforward set of predictions about the role of connectives in pronoun interpretation. While previous research has claimed that connectives have their own focusing properties (Stevenson et al. 1994; Stevenson, Knott, Oberlander & McDonald 2000), this equation would suggest that the relationship between connectives and pronoun interpretation is indirect: Connectives are expected to influence $P(CR)$, which Equation (9) would then predict should affect pronoun interpretation in the manner just described. Consider a connective like *because*, for instance, which disambiguates the ensuing coherence relation to be Explanation. The model would thus predict that, upon hearing *because*, the comprehender's bias for $P(CR = \textit{Explanation})$ will essentially become 1, and thus the probability of $P(\textit{pronoun} = \textit{referent})$ will become $P(\textit{pronoun} = \textit{referent} \mid CR = \textit{Explanation})$. Kehler et al. (2008) presented a study that shows that the next-mention biases for passages with *because* are the same as those when no connective is present, when only the passages related by Explanation in the latter condition are analyzed. As such, this data suggests that it is not the connective that shifts the bias toward particular entities, but the underlying coherence relation (pace Stevenson et al. 2000).

Here we pursue a different test of the predictions of Equation (9). As reported in Rohde, Kehler & Elman (2007) (see also Rohde & Kehler (in prep)), we re-ran the transfer-of-possession experiment just described except with an added manipulation. The manipulation did not involve changing the stimuli, but instead only the instructions. We asked one group of participants to try to answer the question *What happened next?* when they wrote their continuations. Here we hoped to elicit mostly Occasion relations, which we know from the previous experiment to be Goal biased. We asked a second group of participants to instead answer the question *Why?* in their continuations. Here we hoped to get mostly Explanation relations, which the previous experiment demonstrated to be Source biased. The question, then,

is whether this change in the instructions will change the distribution of pronoun assignments, even though the stimuli themselves are identical. Clearly a heuristic approach predicts that it will not, whereas the coherence-driven analysis captured by Equation (9) predicts that it will.

We first ask whether the manipulation succeeded in yielding a different distribution of coherence relations across the two conditions. It did indeed, as can be seen in Table 3. The *What happened next?* condition yielded mostly Occasions, although

	What next?	*Why?*
Occasion	.71	.01
Explanation	.01	.91
Elaboration	.05	.08
Violated Exp	.08	.01
Result	.05	0

Table 3 Distribution of Coherence Relations across Conditions (Perfective Condition)

there were other relations as well (most of which are compatible with the forward movement of time). The *Why?* condition yielded a high proportion of Explanation relations, as anticipated. Thus, $P(CR)$ was successfully shifted. Now what about the pronoun biases $P(\textit{pronoun} = \textit{referent} \mid CR)$? Since the stimuli were the same across conditions, there is no reason to predict a difference in the pronoun interpretation biases when conditioned on coherence relations. As can be seen in Table 4, this was largely the case: The biases found in the previous experiment and the current one were highly consistent.[6]

The instruction manipulation therefore succeeded in shifting $P(CR)$ while keeping $P(\textit{pronoun} = \textit{referent}|CR)$ relatively constant. We now ask whether the shift in $P(CR)$ created a shift in the distribution of pronoun interpretations $P(\textit{pronoun} = \textit{referent})$, as predicted by Equation (9). As can be seen in Table 5, it did: The *Why?* condition yielded significantly more interpretations to the Source referent than the *What happened next?* condition.

It is difficult to see how this result can be reconciled with any theory based primarily on morphosyntactic biases, because the stimuli themselves were identical between conditions. All that changed were the participants' expectations regarding where the discourse was going in terms of coherence. Crucially, note that participants *could* have seen the pronoun prompt, used a morphosyntactic heuristic (subject

6 ANOVAs and logistic regressions confirm that the differences are non-significant.

	$P(pronoun = Source \mid CR)$ (Original Experiment)	$P(pronoun = Source \mid CR)$ (Instruction Manipulation)
Elaboration	.98	1.00
Explanation	.80	.82
Violated Exp	.76	.74
Occasion	.18	.27
Result	.08	.09

Table 4 Biases Conditioned on Coherence Relations (Perfective Condition)

What next?	*Why?*
.34	.82

Table 5 Proportion of Source Interpretations (Instructions Manipulation; Perfective Condition)

assignment, parallel grammatical role, etc.) to fix on a referent (e.g., the subject), and then complete the passage in a way that satisfied the instructions. This is always possible, as the biases toward particular referents associated with coherence relations are merely probabilistic – one can always answer either *What happened next?* or *Why?* coherently while assigning the pronoun to either potential referent. But this is not what participants did. Instead, their choices about what referent to assign the pronoun to were intertwined with their expectations about the direction the discourse was taking in terms of coherence.

At this point, it would appear that the coherence-relation based approach is winning. There can be no doubt that coherence relations influence the assignments of referents to pronouns. But there is another set of data that we have yet to account for. When Stevenson et al. (1994) conducted their study, they also included a no-pronoun prompt condition (10b) alongside their pronoun prompt condition (10a):

(10) a. John passed the comic to Bill. He _____

b. John passed the comic to Bill. _____

As we have previously discussed, Stevenson et al. found an essentially 50/50 interpretation preference for (10a). But for (10b) – where participants not only decide who to refer to first, but also the form of reference to use – they found a strong asymmetry. Specifically, across their context types, they found a strong bias toward

using a pronoun when participants referred to the previous subject, and a strong bias toward using a name when they referred to a non-subject. Stevenson et al. did not break down the statistics by context type, however Arnold (2001) conducted a similar study with transfer-of-possession contexts and found that, for Source-Goal sentences, 76% of the references to the subject (Source) were pronominalized whereas only 20% of the references to the non-subject (Goal) were.[7] This result might seem paradoxical at first: If speakers have a clear preference toward using pronouns to refer to the previous subject and names for non-subjects, why is their pronoun interpretation bias 50/50 in such contexts?

As pointed out in Kehler et al. (2008), one sees that there is nothing paradoxical about the situation when the relationship between interpretation and production is cast using Bayes' Rule:

(11)

$$P(\textit{referent} \mid \textit{pronoun}) = \frac{P(\textit{pronoun} \mid \textit{referent})P(\textit{referent})}{\sum_{\textit{referent} \in \textit{referents}} p(\textit{pronoun} \mid \textit{referent})\ p(\textit{referent})}$$

The term $P(\textit{referent} \mid \textit{pronoun})$ represents the interpreter's problem: He knows that a pronoun was used, but does not know what the intended referent is. On the other hand, the term $P(\textit{pronoun} \mid \textit{referent})$ represents the speaker's problem: She knows what the referent is, but has to decide what form of reference to use. Equation (11) tells us that these biases are not mirror images of each other: Instead, they are related by the prior $P(\textit{referent})$, the probability that a particular referent will get mentioned again, regardless of referring expression.[8] In this model, therefore, a comprehender's interpretation bias thus relies jointly on his estimates of the likelihood that a particular referent will be mentioned next (regardless of form of reference) and of the likelihood that the speaker would have chosen a pronoun (instead of another form of reference) to refer to that referent. In Arnold's data, even though $P(\textit{pronoun} \mid \textit{Goal})$ was only 20%, $P(\textit{Goal})$ was a whopping 86%. Combine these values with the probability of pronominalization in the denominator (28%), and Equation (11) actually predicts a 61% interpretation bias toward the Goal for her stimuli. Hence, there is nothing inconsistent about having an interpretation

7 Miltsakaki (2007) reports on a study of Greek reference that similarly revealed a disconnect between what entities were most likely to be referred to next and the choice of referring expression (in this case, null vs. strong pronoun).

8 The denominator represents $P(\textit{pronoun})$, which can be computed by summing the numerator over all possible referents that are compatible with the pronoun. We ignore this term in our discussion since it contributes a constant factor, having the effect of normalizing the probabilities over all possible referents to 1.

bias toward a non-subject referent despite a strong bias against pronominalizing non-subjects, assuming a suitably large next-mention bias toward the non-subject.

Equation (11) predicts that if production biases favor pronominalization of the subject, we should see an overlaid subject bias in interpretation as well – that is, the interpretation bias $P(Subject \mid pronoun)$ should be higher than the next-mention bias $P(Subject)$. This is precisely what Stevenson et al. found: Across their context types, there was a greater percentage of first mentions to the previous subject when a pronoun prompt was provided (10a) than when it was not (10b). Whereas Stevenson et al. took this as evidence of an interpretation bias toward the subject, equation (11) casts it instead as a production bias.

Whereas our previous work described thus far (Rohde et al. 2006, 2007) always examined coherence-driven biases utilizing contexts with pronoun prompts, we realized that if there is an overlaid subject bias for pronoun interpretation that sits outside of coherence-driven next-mention biases, then form of reference is also predicted to impact expectations about ensuing coherence relations. To see this, consider the equation shown in (12), which models the probability of the ensuing coherence relation based on the conditional probability of a coherence relation given that a particular referent has been mentioned first:

(12)
$$P(CR) = \sum_{referent \in referents} P(referent) * P(CR \mid referent)$$

Again, the next-mention expectation $P(referent)$ is the probability that we estimate from the first mentions in the no-pronoun condition in (10b). If pronouns contribute an overlaid subject bias, then $P(referent)$ is expected to be higher if a pronoun is encountered, even when the pronoun is ambiguous. From this, two predictions result: (i) that there will be more first mentions to the subject in (10a) than (10b) (as Stevenson et al. found), and (ii) that this shift will lead to more subject-biased coherence relations in (10a) than (10b). To test these, we conducted a passage completion experiment using Source-Goal transfer-of-possession contexts like (10a-b) (Rohde & Kehler 2008, in prep).

The results supported the predictions. First, as we expected, previous results concerning the effect of prompt type on next-mention biases were replicated: Whereas 84% of the first mentions were to the Goal in the no-pronoun condition, only 48% of them were to the Goal in the pronoun condition. The distribution of coherence relations is shown in Table 6. To simplify the analysis, we compared the total number of Elaboration and Explanations, which we know to be strongly Source-biased, to the total number of Occasion and Result relations, which we know to be Goal-biased. In the no-pronoun condition, 26% of the relations were Elaborations or

	No-Pronoun Condition	Pronoun Condition
Elaboration	.06	.20
Explanation	.20	.28
Occasion	.36	.28
Result	.13	.05
Violated Exp	.18	.14

Table 6 Distribution of Coherence Relations by Prompt Type

Explanations, whereas 49% were Occasions or Results. As predicted, this distribution shifts towards the Source-biased relations in the pronoun condition, with 48% of the relations being Elaboration or Explanation, and only 33% being Occasion or Result.

As a sanity check, one might ask whether the statistics for the pronoun condition align with those of the no-pronoun condition in those cases in which participants used a pronoun to mention the first referent. The answer is yes: The combined percentage of Elaborations and Explanations in that subset of the no-pronoun condition was 47% (vs. 48% in the Pronoun condition) and the combined percentage of Occasion and Result relations was 33% (vs. 33% in the Pronoun condition). The lack of a statistically significant difference confirms that participants display the same biases regardless of whether a pronoun is provided by the prompt or they choose to use a pronoun themselves.

The results of this experiment show that pronoun interpretation is not simply a by-product of coherence establishment processes along the lines of what Hobbs envisaged. Instead, the dependency between coherence and coreference is bidirectional. That is, the mere occurrence of a pronoun – even one that is ambiguous among multiple possible referents – changes comprehenders' expectations about the direction in which the discourse is going with respect to coherence. This effect is not predicted by a Hobbsian analysis, in which pronouns are merely free variables that rely solely on coherence establishment and other inferential processes for their interpretation, since under this approach the appearance of an ambiguous pronoun in contexts such as these should essentially be inert. Indeed, pronouns are *not* inert, contributing their own linguistic biases to the interpretation process.

To recap, the picture that has emerged is one in which pronoun interpretation is driven by the combination of two sets of biases, as captured by the Bayesian formulation in (11), repeated below as (13):

(13)

$$P(referent \mid pronoun) = \frac{P(pronoun \mid referent)P(referent)}{\sum_{referent \in referents} p(pronoun \mid referent)\ p(referent)}$$

The question now is what factors condition the two terms in the numerator. In Kehler et al. (2008), we noted that all of the data we have considered thus far – including that from Stevenson et al., Arnold, and our own – is "consistent with a scenario in which grammatical or information structural factors (subjecthood, topichood) play a greater role in conditioning $P(pronoun \mid referent)$ and coherence-driven expectations play a greater role in conditioning $P(referent)$" (p. 40). According to this picture, the coherence-driven biases we have documented only affect pronoun interpretation indirectly, by determining the likelihood that a particular referent is mentioned next. The decision to pronominalize, on the other hand, is not expected to be sensitive to those same factors, instead being driven by factors such as the grammatical role and/or topichood status of the referent. This hypothesis results in a surprising prediction: That the rate of pronominalization toward particular referents should not be sensitive to the ultimate interpretation biases towards those referents. This prediction runs directly counter to the common assumption that speakers are more likely to pronominalize a mention of a referent in those cases in which their addressees would be expected to assign the pronoun to that same referent (Arnold 2001; inter alia).

As unintuitive as it may seem, a series of recent studies have supported this prediction. These studies have been conducted using both gender-unambiguous and gender-ambiguous contexts. For instance, Ferretti, Rohde, Kehler & Crutchley (2009)[9] reports on a completion study using the previously-discussed aspect manipulation (perfective vs. imperfective) with no-pronoun prompts in a gender-unambiguous context:

(14) a. Sue handed a timecard to Fred. _____

b. Sue was handing a timecard to Fred. _____

As expected, the percentage of references to the Goal was significantly greater in the perfective contexts than the imperfective ones; this held for all references as well as for the subset of cases in which the participants referred to the first-mentioned participant with a pronoun. This result establishes that, as in previous experiments, the two conditions differ with respect to strength of the next-mention bias toward the Goal. However, this difference did not affect the rate of pronominalization for each

9 See also Rohde (2008), Experiment VII.

referent. While there were significantly more pronouns used to refer to the subject (76.2%) than the non-subject (52.1%), pronominalization did not differ by aspect, nor was there any interaction between referent position and aspect.[10]

Other studies that tested these predictions used so-called IMPLICIT CAUSALITY (IC) verbs, which is undoubtedly the most well-studied verb class in the psycholinguistics of pronoun interpretation literature since the seminal papers of Caramazza and colleagues in the 1970s (Garvey & Caramazza 1974; Caramazza, Grober, Garvey & Yates 1977; Brown & Fish 1983; Au 1986; McKoon, Greene & Ratcliff 1993; Koornneef & van Berkum 2006; Kehler et al. 2008; inter alia). If participants are presented with a prompt like (15a),

(15) a. Amanda amazed Brittany because she _____

b. Amanda detested Brittany because she _____

the large majority of completions will point to Amanda as the pronominal referent. After all, Amanda must be amazing, and hence one expects to hear why. Because causality is imputed to the subject, verbs like *amaze* are called SUBJECT-BIASED IC VERBS. If you give participants a prompt like (15b), on the other hand, the large majority of completions will point to Brittany as the pronominal referent. After all, Brittany must be detestable, and hence one expects to hear why. Because causality is imputed to the object, verbs like *detest* are called OBJECT-BIASED IC VERBS.

The existence of IC biases has been replicated repeatedly, and is hence one of the bedrock results in the field. Kehler et al. (2008) also showed the IC verbs are associated with a second type of strong bias. Whereas most previous studies examined pronoun biases using prompts with *because*-clauses as in (15a-b) above, Kehler et al. used full-stop prompts as in (16), and had annotators categorize the coherence relations in the completed passages.

(16) a. John amazed Mary. _____

b. John detested Mary. _____

c. John saw Mary. _____

10 In contrast to this result, Arnold (2001) found an effect of thematic role on rate of pronominalization in her comparison of Goal-Source and Source-Goal transfer-of-possession contexts, in which references to the Goal were pronominalized more often than references to the Source. As pointed out by Fukumura & van Gompel (2010), however, her two context types differed not only in the order of the thematic role fillers but in a number of other relevant respects, including the verbs themselves (which are known to vary in the strength of their biases even within the same category). The difference in rate of pronominalization between Goals and Sources, while significant, was relatively small, especially in comparison to the effect of grammatical role.

The results showed that participants completed passages (16a-b) using Explanation relations approximately 60% of the time, compared to 24% for a control group of non-IC contexts (16c). This result accords with intuitions: Upon hearing *John amazed Mary*, it seems likely that the addressee will wonder *Why?*, and thus expect to hear an answer to this question. On the other hand, when someone says *John saw Mary*, it seems less likely that the addressee will wonder *Why?*, and instead expect an answer to the question *What happened next?*, for example.

The diametrically-opposed next-mention biases for IC verbs make them a good choice for examining the effects of interpretation biases on pronoun production. Rohde (2008: Experiment V) reported on a passage completion study that, like the study with transfer-of-possession contexts just described, used gender-unambiguous contexts and no-pronoun prompts of the sort shown in (16a-c). Again, the rate of pronominalization varied by the grammatical role of the referent (98% pronominal mentions for subjects vs. 77% for non-subjects). However, there was no interaction between grammatical role and context type – references to the object were no more likely to be pronominalized following an object-biased IC verb than following a subject-biased IC verb.

Fukumura & van Gompel (2010) similarly conducted two experiments using IC contexts that were gender-unambiguous. In the first experiment, they varied between contexts that employed subject-biased and object-biased IC verbs:

(17) a. Gary scared Anna after the long discussion ended in a row. This was because _____

b. Gary feared Anna after the long discussion ended in a row. This was because _____

While there was a greater rate of pronominalization to the previous subject than the object, they also found that the IC bias made no difference (subjects: 76% for subject-biased vs. 74% for object-biased; objects: 61% for subject-biased vs. 60% for object-biased). Their second experiment used only subject-biased IC contexts, and instead varied the connective between *because* and *so*:

(18) a. Gary scared Anna after the long discussion ended in a row because _____

b. Gary feared Anna after the long discussion ended in a row so _____

It is known that biases for such contexts flip from the subject with *because* toward the object with *so* (Stevenson et al. 1994, 2000; Crinean & Garnham 2006). As

predicted, while the context manipulation affected who was mentioned next in the continuation, it did not affect rate of pronominalization; only grammatical role mattered.

One might argue that the foregoing studies do not fully test the predictions of the analysis, however, since in each case the contexts mention two opposite-gender individuals, and hence the subsequent pronoun is never ambiguous. That is, perhaps speakers only take interpretation biases into account when the comprehender's ability to resolve the pronoun to the correct referent is actually at stake. This turns out not to be the case. Rohde & Kehler (2013)[11] describes a passage completion study that used the stimuli from the earlier study except that the competing referents were of the same gender:

(19) a. John amazed Bill. _____

b. John detested Bill. _____

c. John saw Bill. _____

Once again, whereas next-mention biases varied between the contexts, the rate of pronominalization was influenced only by the grammatical role of the referent, with no interaction between grammatical role and context type.[12]

The studies just described therefore support the idea that choice of referring expression is determined by grammatical role. The rate of pronominalization to the subject is consistently greater than to the non-subject, whereas the semantic biases that have been established to influence interpretation have no effect. However, a heuristic preference towards grammatical subjects is not the only explanation that is consistent with the production facts. The aforementioned quote from Kehler et al. (see p. 18) distinguished subjecthood biases, which are rooted in syntax, and topichood biases, which are rooted in pragmatics (specifically, information structure). The sentence topic is commonly (albeit informally) characterized as the constituent that expresses what the sentence is about, that is, the entity about which an utterance is primarily intended to expand the addressee's knowledge (Strawson 1964; Kuno 1972; Gundel 1974; Reinhart 1981; Givón 1983; Lambrecht 1994; inter alia). While

11 See also Rohde (2008), Experiment VI.

12 Kaiser (2010) also examined the relationship between pronoun production and interpretation biases in ambiguous contexts, but instead of manipulating verb-driven semantic biases like the experiments just described, she manipulated the syntax and information structure of the context sentence, crossing SVO and clefted variants with subject and object focus. She similarly found a strong dissociation between rate of next mention and rate of pronominalization. Particularly striking was the fact that she found a large next-mention bias toward previously-mentioned alternatives to contrastively-focused constituents (particularly in subject-focus constructions), but these referents were almost never mentioned using a pronoun, nor were they referred to in a pronoun-prompt version of the experiment.

subject position is the canonical place for a topic to appear in English, the two notions cannot be conflated; sometimes a non-subject serves the role of sentence topic instead of the subject (Lambrecht 1994: p. 118). For instance, whereas Amanda may normally be considered more topical than Brittany in the sentence *Amanda amazes Brittany*, the preferred topic of the sentence *No one amazes Brittany* is no doubt Brittany, as the sentence intuitively predicates a property of her rather than *no one*. Although sentence position is important, a number of other factors (referentiality, definiteness, context, etc.) affect the topicality of referents as well.

It is here that Centering Theory comes back into the picture. Recall the aforementioned quote from Grosz et al. (1995):

> ...the use of a pronoun to realize the C_b signals the hearer that the speaker is continuing to talk about the same thing. (GJW 95, p. 214)

Since the C_b is an operationalization of the sentence topic, this principle amounts to a production bias toward pronominalizing the topic (see also Ariel (1990), Gundel, Hedberg & Zacharski (1993), and Lambrecht (1994), for similar claims). If this is correct, then the factors that drive pronominalization primarily concern topicality, with which subjecthood is merely correlated. Under this hypothesis, therefore, one could attribute the declining rates of pronominalization of entities in particular grammatical roles as one moves down the obliqueness hierarchy not to a bias specific to grammatical role, but instead to the fact that the likelihood of being the topic declines along the obliqueness hierarchy. As such, the bias toward pronominalizing the topic might be very strong as Centering would have it; the uncertainty enters with respect to the likelihood that the particular entity being referred to is the topic.

This hypothesis therefore predicts that the rate of pronominalization of a referent will in fact vary with its likelihood of serving as the topic, despite the lack of effect of semantic biases on production. As such, if we had a way of manipulating likelihood of topicality while keeping grammatical role constant, we would nonetheless expect a difference in the rate of pronoun production. An alternation that meets this criterion in English is the variation between active and passive voice. Whereas the context sentences in (20a) and (20b) below describe the same proposition, they differ not only in which entity is considered the default topic (Amanda in 20a and Brittany in 20b), but also with respect to the likelihood that their respective subjects serve as the topic. That is, whereby Amanda is merely the default topic in (20a) outside of a larger context, the promotion of Brittany to the subject position in (20b) by way of a marked syntactic construction constitutes a much stronger signal that she is the topic. Indeed, providing a syntactic means for predicating a property of an entity other than the one that would normally serve as the subject of the active form is commonly considered to be one of the primary functions of the passive (Davison

1984; inter alia).

We therefore conducted an experiment using subject-biased IC contexts in a 2x2 manipulation (Rohde & Kehler 2009, 2013):

(20) a. Amanda amazed Brittany. She _____

b. Brittany was amazed by Amanda. She _____

c. Amanda amazed Brittany. _____

d. Brittany was amazed by Amanda. _____

Examples (20a-d) differ in terms of whether the context sentence is in the active or passive voice (voice manipulation) and whether or not a pronoun is included in the prompt (prompt manipulation). We examine a series of predictions made by the analysis we have developed. First, we predict that the voice manipulation will affect the pronoun interpretation bias. A pure version of the coherence-relation theory (of the sort that Hobbs proposed) would predict no difference between (20a-b), since the manipulation is syntactic and hence does not alter the context in terms of meaning. The model espoused here, however, predicts that the overlaid subject bias for production will lead to fewer interpretations to the causally-implicated referent (Amanda) in the passive (20b) than in the active (20a), since in (20b) this referent has been moved out of subject position. Second, we also predict that there will be more subject references in each pronoun prompt condition (20a-b) than in their respective no-pronoun conditions (20c-d), as has been found repeatedly in previous work. Third, following the results of Rohde & Kehler (2008), we predict that the switch between the active and passive voice should not only influence pronoun interpretations, but also the distribution of coherence relations. In particular, by drawing references away from the causally-implicated referent, the use of the passive (20b) should cause participants to write fewer Explanations than in the active case (20a). This too would be surprising on a purely coherence-based account, since there is no reason to think that a change in syntactic form that keeps semantics constant would have an effect on expectations about coherence relations.

The final prediction is the central one regarding production: That the difference between active and passive will yield a greater number of pronominalizations of the subject in the passive case (20d) as compared to the active case (20c) in the no-pronoun condition. As argued earlier, this prediction arises from the fact that, in English, being the subject of a passive clause is a stronger indicator of being the sentence topic than being the subject of an active clause. A significant difference would stand in stark contrast with the studies we have surveyed to this point, since the rate of pronominalization across conditions for referents in the same grammatical role

	Pronoun	No Pronoun
Active	.77	.59
Passive	.42	.76

Table 7 Proportion of Next Mentions of the Causally-Implicated Referent

has always been constant. The difference here is that topicality is being manipulated, rather than the sorts of semantic biases that have been shown to affect interpretation.

All of these predictions were borne out. First, as shown in Table 7, the voice manipulation did affect pronoun interpretation biases: References to the causally-implicated referent – i.e., Amanda in (20a-b) – were significantly higher in the active condition (ex. 20a, 77%) than in the passive condition (ex. 20b, 42%). That is, putting Brittany in subject position actually resulted in her being the preferred referent, even though she is the bias-incongruent entity with respect to IC.[13] Also, as expected, the percentage of first mentions of the subject in the active condition was higher when a pronoun prompt was given (ex. 20a, 77%) than when no pronoun was given (ex. 20c, 59%).[14]

	Pronoun	No Pronoun
Active	.75	.60
Passive	.52	.72

Table 8 Proportion of Explanation Relations

Next, as predicted, the voice manipulation significantly affected the distribution of coherence relations. As shown in Table 8, in the pronoun condition there

13 This result is consistent with that of Caramazza & Gupta (1979), who found an effect of passivization in IC contexts using a timed comprehension task. See also Kaiser, Li & Holsinger (2011) for a complementary result for passivization with Agent-Patient verbs.

14 Interestingly, the percentage of first mentions of the causally-implicated referent in the no-pronoun condition was greater in the passive case (when the referent appeared as the surface object of the *by*-phrase; 76%) than in the active one (59%). We attribute this to the fact that *by*-phrases in English passives are optional, and hence their appearance in the context sentence might have driven participants to feel the need to re-mention the referent in order to justify its inclusion in the story. This interpretation follows that proposed by Arnold (2001), who found the same effect for contexts with Goal-Source transfer-of-possession verbs (*Sue received a book from Mary*), where the non-subject referent is similarly found in an optional PP. Importantly, the large majority of these mentions used names, and hence the pattern found for the no-pronoun condition was not similarly exhibited in the pronoun condition.

were more Explanation relations in the active case (75%) than in the passive case (52%). This result is again surprising on a purely semantic theory of coherence establishment, since the propositions expressed by (20a-b) are invariant between the two syntactic manifestations. Here, the shift in coherence relations is mediated by the shift in the pronominal reference distribution: Since passivization results in a greater number of pronoun interpretations to the non-causally-implicated referent, participants were less likely to continue the passage with an Explanation relation.[15]

As predicted, the effect of passivization on referential form production was also confirmed. As shown in Table 9, a greater percentage of references to the previous subject were pronominalized in the passive condition (87%) than in the active condition (62%). This result is expected on a theory that posits that topics

	Surface Subject	Surface Non-Subject
Active	.62	.24
Passive	.87	.23

Table 9 Proportion of Pronominalized References

should be pronominalized (e.g., Centering), since appearing as the subject of a passive construction is a stronger indicator of being the sentence topic than being the subject of an active voice clause. This result therefore suggests that the production biases are conditioned on information-structural factors rather than grammatical ones, as not all grammatical subjects are considered equal. There was no difference statistically in the rate of pronominalization of the non-subjects in each case (24% for actives and 23% for passives).

Finally, we can also use the data collected here to test what might be considered to be the core prediction of the Bayesian analysis, specifically that the relationship between pronoun production and interpretation biases is as predicted by equation (13). To do this, we estimate the values for the terms on the right-hand side of equation (13) from the data collected in the no-pronoun prompt conditions, which will yield predictions for $p(\textit{referent} \mid \textit{pronoun})$ on the left-hand side. If the Bayesian characterization is correct, these predicted biases should be highly correlated with the actual interpretation biases that are estimated directly in the pronoun-prompt conditions of the experiment. The predicted and actual interpretation probabilities are listed in Table 10, broken down by voice. There is a strong alignment between

15 In the no-pronoun condition, there were more Explanation relations in the passive case (20d) than in the active case (20c), even though the causally-implicated referent appeared in the by-phrase. This follows from the fact that there were more first-mentions of the causally-implicated referent in the passive condition, as explained in footnote 14.

these values in each context condition; linear models reveal that the predicted and actual values are in fact highly correlated, thus confirming the predictions of the model.

	Predicted	Actual
Active	0.81	0.74
Passive	0.59	0.60

Table 10 Predicted and Actual Biases for Pronominal Reference to the Subject (Voice Manipulation)

To sum, the results confirm our central prediction about pronoun production, in that participants chose to use a pronoun more often when referring back to the subject of a passive voice clause than the subject of an active. This result contrasts with previous results in which manipulations outside of grammatical role (specifically, semantic bias) had no effect on rate of pronominalization. As expected, we also find a difference in interpretation bias between the conditions, whereby there were more references to the bias-inconsistent referent when it occupied the subject position rather than the non-subject position. Further, this change in preferred referent for the pronoun cascaded to create a shift in the distribution of coherence relations – specifically, away from Explanation relations – demonstrating again that the relationship between coherence and coreference is bidirectional. Finally, the data confirmed the predictions of the Bayesian analysis, whereby interpretation biases estimated from the no-pronoun condition were highly correlated with the actual biases measured in the pronoun condition.

4 Discussion

We have compared two classic theories of pronoun interpretation that have sought to specify the relationship between pronoun use and discourse coherence; theories that make seemingly irreconcilable claims. According to Hobbs (1979, 1990), pronoun interpretation is not governed by an independent mechanism, but instead comes about as a by-product of utilizing world knowledge during the inferential establishment of coherence relations, without regard to the grammatical form and information structure of utterances. According to Centering Theory (Grosz et al. 1995; inter alia), on the other hand, pronoun interpretation is predominantly determined by information structural relationships within and between utterances (e.g., topic transitions) and the grammatical roles occupied by potential referents, without appeal to world knowledge or the establishment of informational coherence relations.

In this paper, we have argued – based on the results of a series of psycholinguistic experiments – for an analysis that reconciles these diverse approaches. The reconciliation comes by way of a simple, Bayesian probabilistic model in which the interpretation bias $P(\textit{referent} \mid \textit{pronoun})$ is determined by:

- Expectations that comprehenders have about what Hobbsian coherence relations will ensue, which in turn condition top-down expectations about referent next mention (regardless of the referring expression used), as modeled by the prior $P(\textit{referent})$, and
- Centering-style constraints on pronoun production, which provide bottom-up evidence about the topichood status of referents that are specific to the speaker's decision to use a pronoun, as modeled by the likelihood $P(\textit{pronoun} \mid \textit{referent})$.

The analysis comes a long way from the kind of SMASH approach found in the literature, whereby reference resolution is the result of an often complex search-and-match procedure that is triggered when a pronoun is encountered. Indeed, the analysis fits well with what has come to be a modern view of language interpretation in psycholinguistics, in which interpretation is not something that initiates when linguistic material is encountered, but is instead what occurs when top-down expectations about the ensuing message come into contact with the bottom-up linguistic evidence.

Whereas both the Hobbsian coherence analysis and Centering emerge as contributing important insights, note that neither account as stated turns out to be adequate in and of itself as a theory of pronoun interpretation. Our results suggest that Hobbsian coherence is crucial in generating expectations about likelihood of entity mention, but that it achieves its influence on pronoun interpretation only indirectly. Indeed, the data indicate that the idea that pronouns merely function as unbound variables in the way Hobbs envisages is untenable. Centering, likewise, contributes correct predictions to the problem of pronoun *production*, and hence, it also achieves its influence on pronoun interpretation only indirectly. Notably, we have had no reason to appeal to any existing Centering-based algorithms for pronoun *interpretation* (Brennan et al. 1987; Walker et al. 1994; Strube & Hahn 1999; inter alia). These analyses use transition hierarchies to rank pronoun assignments, where the transition type is determined solely by the identities of the C_b and C_f. It is hard to see how any analysis that does not incorporate pragmatically-driven next-mention biases could account for the range of data addressed here.

Historically, proponents of the Centering and Hobbsian approaches have appealed to different types of discourses to motivate their theories. Whereas the evidence argued to support the Hobbsian approach often consists of short, semantically-

biased passages (4), the data provided in support of Centering are typically longer, multi-utterance discourses in which a topic has been more clearly established through repeated mention in subject position (Grosz et al. 1995; Brennan et al. 1987; inter alia). The analysis presented here explains the motivation for this difference, as it predicts that the more certain the hearer is about the identity of the current topic (which will drive up his estimate for $P(\textit{pronoun} \mid \textit{referent})$), the less that coherence-driven next-mention biases will be able to move the interpretation preference for a subsequent pronoun away from that entity. We would thus expect to see a greater pronoun bias toward the subject in passage (22) than the original two-sentence version repeated below in (21), in so far as the repeated mention of John increases the likelihood that he is the topic.

(21) John passed a comic to Bill.

He _____

(22) John walked into the room.

He saw Bill sitting around with nothing to do.

He passed a comic to him.

He _____

Whereas this prediction seems intuitively correct to us, it awaits experimental confirmation.

In keeping with the goals laid out in the Introduction, our analysis offers insight into why we see evidence for so-called 'preferences' in pronoun interpretation, and why different preferences seem to be determinative in different contextual circumstances. The analysis explains the paradox whereby choice of syntactic form affects interpretation biases even when the propositional content of the context is unchanged (as in minimal pairs that vary between passive and active voice, e.g. 1a-b and 20a-b), while at the same time we find other cases in which syntactic form is not varied but semantic content is (as in examples 4a-b and 19a-c). Whereas other analyses have posited both syntactic and semantic influences on pronoun interpretation, these influences are often treated as an unstructured collection, without higher-level causal structure spelling out the relationships among them. The analysis put forth here is perhaps unique in positing a particular set of hierarchical relationships among the different factors, with the result that many factors that are considered primary in other theories (e.g., subject biases, thematic role biases, and so forth) are rendered derivative in ours. While it is true that statistical analysis of these more superficial

factors may reveal significant effects, we contend that these factors are epiphenomena of the true underlying causes of pronoun behavior.

Whereas our results are incompatible with heuristic approaches, they are also inconsistent with analyses that equate pronoun interpretation biases with a notion of accessibility and/or probability of next mention, such as Arnold's Expectancy Hypothesis (Arnold 2001, 2010). In Arnold's framework, a variety of biases – grammatical subjecthood, parallel grammatical role, thematic role, topichood, and others – all work as a collection to confer accessibility to referents. This approach leaves us without an explanation, however, for why pronoun interpretation biases (as measured by pronoun-prompt conditions in completion studies) can vary dramatically from next-mention biases (as measured by no-pronoun conditions). Indeed, our data suggest that it is *always* felicitous to pronominalize a reference to the subject of the previous sentence, *no matter how far the interpretation bias points away from that referent*. Arnold (2010) suggests that "the likelihood of re-mention could be construed as one measure of topicality", but our data indicates that topicality and likelihood of re-mention play distinct and non-conflatable roles: Likelihood of re-mention affects interpretation but not production, whereas topicality is a central criterion for production. In her discussion of what she calls the COMMUNICATIVE ROLE OF REFERRING HYPOTHESIS, she observes that "the traditional view is that speakers choose referential forms on the basis of interpretability" (p. 197). While this is certainly true to an extent, our data indicate that there are limits: Speakers will pronominalize mentions in some cases in which the referent is not the one preferred by interpretation biases, and will likewise use names to refer to referents which would have been the preferred referent had a pronoun been used.

Importantly, our analysis should not be construed as claiming that no factors other than those we have discussed influence pronominal reference. Indeed, in Kehler et al. (2008) we argued that information-structural constraints that govern the placement of accent in English explain the evidence for another purported preference, the grammatical role parallelism bias. Recall that evidence for the bias comes from examples like (1c), repeated in (23a), but that such evidence is mysteriously absent in passages like (1a), repeated in (23b).

(23) a. Mitt narrowly defeated Rick, and Newt absolutely trounced him. [him = Rick]

b. Mitt narrowly defeated Rick, and the press promptly followed him to the next primary state. [him = Mitt]

Kehler et al. (2008) argue that in cases of RESEMBLANCE coherence relations – such as the Parallel relation in (23a) – information-structural factors conspire to necessitate that any referring expression that is not coreferent with its semantically

parallel constituent in the first clause (i.e., *Rick* in 23a) must receive accent. As such, while it is possible for a pronoun to corefer with a non-parallel antecedent in such relations, it must receive accent to do so, as shown in (24a):

(24) a. Mitt narrowly defeated Rick, and Newt absolutely trounced HIM/#him. [him = Mitt]

b. Mitt narrowly defeated Rick, and Newt absolutely trounced MITT/#Mitt.

Importantly, the requirement for accent applies whether or not the referring expression is a pronoun (24b), demonstrating that the constraint is independent of pronominalization (Akmajian & Jackendoff 1970; Venditti, Stone, Nanda & Tepper 2002; de Hoop 2004; inter alia). Once this constraint is acknowledged, it can be seen that there is no work left for an additional parallel grammatical role preference to do. On the one hand, the preference is superfluous when a Resemblance relation is operative, since the rules of accent placement alone necessitate that an unaccented pronoun will refer to a parallel referent. On the other hand, these same constraints are predicted not to apply in cases of non-Resemblance, a prediction which examples like (23b) confirm.[16] The picture that results is one in which the underlying basis of the various biases that affect pronoun interpretation – those driving next-mention expectations (coherence-driven biases), those driving the decision to pronominalize (topichood biases), and those governing when a pronoun may or may not remain unaccented (constraints on accent placement) – while different, are all fundamentally pragmatic.

Our arguments have focused exclusively on the results of passage completion studies. Although the passage completion technique does not provide on-line measurements of interpretation difficulty, it has important advantages over other methodologies for our purposes. For one, it gives us a way to obtain estimates of fine-grained probabilistic biases in a fairly straightforward way. This ability is not only crucial with respect to measuring pronoun interpretation and production biases, but also for estimating expectations about ensuing coherence relations: It is these estimates that have allowed us to make progress on developing a coherence-driven theory of pronoun interpretation without having a fully predictive theory of coherence establishment in hand. Second, the methodology allows us to measure the contribution of a pronoun against an appropriate baseline, as provided by the next-mention expectations measured in the no-pronoun conditions. In our view, there is a

16 Pronouns in non-Resemblance relations may still preferentially refer to entities that occupy grammatically-parallel positions, of course. The point is that such reference will be determined by other factors. For instance, subject/topic biases will point to the grammatically-parallel referent for subject pronouns. Likewise for cases in which semantic biases suggest a referent that happens to be in a grammatically-parallel position.

persistent methodological problem in the literature with respect to stimuli creation, whereby a set of stimuli are constructed that seem intuitively unobjectionable (i.e., 'pragmatically unbiased'), and the biases measured for them are compared against a hypothetical uniform distribution as a baseline (i.e., a 50/50 distribution if there are two competing referents). This may seem reasonable on a heuristic account of pronoun interpretation strategies – a greater than 50% bias to the subject amounts to a subject assignment strategy, and so forth – but it makes little sense on the current account. On this account, only one of the two biases that feed pronoun interpretation is specific to pronouns (the production bias); the other – the prior probability that the referent will be mentioned next – holds regardless of referential form. As such, if one wants to measure the contribution that is specific to pronouns, one must compare the resulting distribution against the distribution found in the no-pronoun condition. The 50/50 point along the spectrum of interpretation biases is as arbitrary a baseline as any other.

This insight helps to clarify what may initially appear to be a self-contradictory set of interpretation facts. That is, at first blush it may seem that pronoun interpretation biases go in many different directions – toward the subject (as in Goal-Source transfer-of-possession contexts, subject-biased IC contexts, and many others), closer to 50/50 (as in Source-Goal transfer-of-possession contexts), or even toward a non-subject (as in object-biased IC contexts). But it would be a mistake to characterize the bias associated with pronouns as being mixed. The reason for this is that in all passage completion studies we are aware of that compare pronoun and no-pronoun prompt data, the inclusion of a pronoun prompt consistently shows a subject bias, in that it *always* pulls the distribution of references to the subject as compared to when a pronoun is not provided. In this sense, a pronoun can display a subject bias even when the resulting interpretation bias is toward a non-subject. The only part of the equation that shows considerable variance across context types is the one which captures next-mention expectations, which again are independent of referential form. Many methodologies provide no straightforward way of teasing apart these fundamentally different and interacting biases, which we believe has obfuscated our understanding of the mechanisms that underlie pronoun interpretation.

That having been said, we have carried out a number of studies using on-line methodologies that provide support for the claim that comprehenders utilize the kinds of biases that we have argued for in this paper. Ferretti et al. (2009), for example, conducted an ERP experiment which found that comprehenders show differential ERP responses when resolving pronouns in contexts containing imperfective and perfective transfer-of-possession verbs, as a result of the differing degree of centrality of the Goal for the two aspects. Rohde & Kehler (in prep) describes a reading time study modeled after the passage completion experiment from Rohde et al. (2007) described earlier, which showed that clauses containing a pronominal reference

to the Goal in transfer-of-possession contexts were read more slowly when the participants were expecting the second sentence to answer to the question *Why?* (which signals an Explanation relation, which is biased to the Source) than when they were expecting an answer to the question *What happened next?* (which signals an Occasion relation, which is biased to the Goal). Finally, Rohde, Levy & Kehler (2011) showed that the expectations that object-biased IC verbs create for an ensuing Explanation, along with the next-mention biases toward their direct objects, influence syntactic attachment biases in online relative clause processing.

Whereas the Hobbsian coherence and Centering models have long been thought to be diametrically opposed theories of discourse coherence and its relationship to pronoun use, we believe that the data presented here have shown that they are in fact largely compatible. Both theories have core principles that the evidence supports, at the expense of requiring that other, less significant aspects be abandoned. It is worth noting that this reconciliation required that we go beyond these accounts in several important ways. The first is that we had to move from a symbolic to a probabilistic framework. We find that no rule-based algorithm will be sufficient for explaining the data presented here, nor is there any obvious way to integrate the theories without appealing to probabilities. The second is that, like Stevenson et al. (1994) and Arnold (2001), the analysis makes crucial appeal to expectations that are brought into the interpretation process rather than merely processing that occurs when a pronoun is encountered. Because reasoning about coherence relations in the Hobbsian framework requires full propositional representations in order to commence, there is no obvious way of accounting for the considerable amount of experimental evidence that pronouns are interpreted incrementally, before the end of the utterance containing them is reached. Further, as argued in Kehler (1997), the classic Centering-based theories of interpretation similarly require, at least in some contexts, that full sentences be processed before the preferred assignments for pronouns can be computed. The model we have presented shows how a coherence-driven theory can be consistent with incremental interpretation, as the information utilized by the model is available to the comprehender at the time the pronoun is encountered. Finally, it is worth stressing the role of experimental manipulations in the collection of this data. Only by conducting experiments do we have a way to estimate biases that can be empirically validated while controlling for other contextual influences. Indeed, we believe that this work provides a useful case study for how experimental work can drive the development of linguistic theories.

References

Akmajian, Adrian & Ray S. Jackendoff. 1970. Coreferentiality and stress. *Linguistic Inquiry* 1(1). 124–126.

Ariel, Mira. 1990. *Accessing noun phrase antecedents*. Routledge.

Arnold, J. E. 2001. The effects of thematic roles on pronoun use and frequency of reference continuation. *Discourse Processes* 31. 137–162.

Arnold, J. E. 2010. How speakers refer: The role of accessibility. *Language and Linguistic Compass* 4. 187–203.

Au, T. K. 1986. A verb is worth a thousand words: The causes and consequences of interpersonal events implicit in language. *Journal of Memory and Language* 25. 104–122.

Brennan, Susan E., Marilyn W. Friedman & Carl J. Pollard. 1987. A centering approach to pronouns. In *Proceedings of the 25th meeting of the association for computational linguistics*, 155–162. Stanford, CA.

Brown, R. & D. Fish. 1983. The psychological causality implicit in language. *Cognition* 14. 237–273.

Caramazza, A., E. Grober, C. Garvey & J. Yates. 1977. Comprehension of anaphoric pronouns. *Journal of Verbal Learning and Verbal Behaviour* 16. 601–609.

Caramazza, A. & S. Gupta. 1979. The roles of topicalization, parallel function and verb semantics in the interpretation of pronouns. *Linguistics* 3. 497–518.

Chambers, G. C. & R. Smyth. 1998. Structural parallelism and discourse coherence: A test of centering theory. *Journal of Memory and Language* 39. 593–608.

Crawley, R., R. Stevenson & D. Kleinman. 1990. The use of heuristic strategies in the interpretation of pronouns. *Journal of Psycholinguistic Research* 4. 245–264.

Crinean, Marcelle & Alan Garnham. 2006. Implicit causality, implicit consequentiality, and semantic roles. *Language and Cognitive Processes* 21(5). 636–648.

Davison, Alice. 1984. Syntactic markedness and the definition of sentence topic. *Language* 60(4). 797–846.

Di Eugenio, Barbara. 1998. Centering in Italian. In Marilyn Walker, Aravind Joshi & Ellen Prince (eds.), *Centering in discourse*, 115–137. Oxford University Press.

Ferretti, Todd, Hannah Rohde, Andrew Kehler & Melanie Crutchley. 2009. Verb aspect, event structure, and coreferential processing. *Journal of Memory and Language* 61. 195–205.

Frederiksen, J. 1981. Understanding anaphora: Rules used by readers in assigning pronominal referents. *Discourse Processes* 4. 323–347.

Fukumura, K. & P. G. van Gompel. 2010. Choosing anaphoric expressions: Do people take into account likelihood of reference? *Journal of Memory and Language* 62. 52–66.

Garvey, C. & A. Caramazza. 1974. Implicit causality in verbs. *Linguistic Inquiry* 5. 459–464.

Gernsbacher, M. A. & D. J. Hargreaves. 1988. Accessing sentence participants: The advantage of first mention. *Journal of Memory and Language* 27. 699–717.

Gernsbacher, M. A., D. J. Hargreaves & M. Beeman. 1989. Building and accessing

clausal representations: the advantage of the first mention versus the advantage of clause recency. *Journal of Memory and Language* 28. 735–755.

Givón, Talmy (ed.). 1983. *Topic continuity in discourse: A quantitative cross-language study*. John Benjamins.

Gordon, P. C., B. J. Grosz & L. A. Gilliom. 1993. Pronouns, names, and the centering of attention in discourse. *Cognitive Science* 17. 311–347.

Grosz, B. J., A. K. Joshi & S. Weinstein. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics* 21. 203–225.

Gundel, Jeanette. 1974. *The role of topic and comment in linguistic theory*. Ph.D. thesis, University of Texas at Austin. Reprinted in Outstanding Dissertations in Linguistics Series, Garland Publishers, 1989.

Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69(2). 274–307.

Hedberg, Nancy. 2010. Centering and noun phrase realization in Kaqchikel Mayan. *Journal of Pragmatics* 42(7). 1829–1841.

Hobbs, J. R. 1979. Coherence and coreference. *Cognitive Science* 3. 67–90.

Hobbs, J. R. 1990. *Literature and cognition*. Stanford, CA: CSLI. Lecture Notes 21.

Hoffman, Beryl. 1998. Word order, information structure, and centering in Turkish. In Marilyn Walker, Aravind Joshi & Ellen Prince (eds.), *Centering in discourse*, 251–271. Oxford University Press.

de Hoop, Helen. 2004. On the interpretation of stressed pronouns. In Reinhard Blutner & Henk Zeevat (eds.), *Optimality theory and pragmatics*, 25–41. New York: Palgrave Macmillan.

Järvikivi, J., R. van Gompel, J. Hyönä & R. Bertram. 2005. Ambiguous pronoun resolution: contrasting the first mention and subject preference accounts. *Psychological Science* 16. 260–264.

Kaiser, Elsi. 2010. Investigating the consequences of focus on the production and comprehension of referring expressions. *International Review of Pragmatics* 2(2). 266–297.

Kaiser, Elsi, David Cheng-Huan Li & Edward Holsinger. 2011. Exploring the lexical and acoustic consequences of referential predictability. In I. Hendrickx, A. Branco, S. Lalitha Devi & R. Mitkov (eds.), *Anaphora processing and applications*, 171–183. Lecture Notes in Artificial Intelligence, Vol. 7099. Heidelberg: Springer.

Kameyama, M. 1986. A property-sharing constraint in centering. In *Proceedings of the 24th annual meeting of the association for computational linguistics*, 200–206.

Kehler, A. 2002. *Coherence, reference, and the theory of grammar*. Stanford, CA: CSLI Publications.

Kehler, A., L. Kertz, H. Rohde & J. Elman. 2008. Coherence and coreference

revisited. *Journal of Semantics* 25. 1–44.

Kehler, Andrew. 1997. Current theories of centering for pronoun interpretation: A critical evaluation. *Computational Linguistics* 23(3). 467–475.

Kehler, Andrew. 2007. Rethinking the SMASH approach to pronoun interpretation. In Jeanette Gundel & Nancy Hedberg (eds.), *Interdisciplinary perspectives on reference processing*. Oxford University Press, New Directions in Cognitive Science Series.

Koornneef, Arnout W. & Jos J. A. van Berkum. 2006. On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye-tracking. *Journal of Memory and Language* 54. 445–465.

Kuno, Susumu. 1972. Functional sentence perspective. *Linguistic Inquiry* 3(3). 269–320.

Lambrecht, K. 1994. *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press.

McKoon, G., S. Greene & R. Ratcliff. 1993. Discourse models, pronoun resolution, and the implicit causality of verbs. *Journal of Experimental Psychology* 19. 1040–1052.

Miltsakaki, Eleni. 2007. A rethink of the relationship between salience and anaphora resolution. In A. Branco (ed.), *Proceedings of the 6th discourse anaphora and anaphor resolution colloquium*, 91–96. Lago, Portugal.

Rambow, Owen. 1993. Pragmatic aspects of scrambling and topicalization in german. In *Working notes of the workshop on centering theory in naturally occurring discourse*. University of Pennsylvania.

Reinhart, Tanya. 1981. Pragmatics and linguistics: An analysis of sentence topics. *Philosophica* 27(1). 53–94.

Rohde, H. 2008. *Coherence-driven effects in sentence and discourse processing*. Ph.D. thesis, UC San Diego.

Rohde, H. & A. Kehler. 2008. The bidirectional influence between coherence establishment and pronoun interpretation. In *Poster at the 21st Annual CUNY Conference on Sentence Processing*. University of North Carolina at Chapel Hill.

Rohde, H., A. Kehler & J. Elman. 2006. Event structure and discourse coherence biases in pronoun interpretation. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society*. Vancouver, July 26-29, 2006.

Rohde, H., A. Kehler & J. Elman. 2007. Pronoun interpretation as a side effect of discourse coherence. In *Proceedings of the 29th Annual Conference of the Cognitive Science Society*. Nashville, TN, August 1-4, 2007.

Rohde, Hannah & Andrew Kehler. 2009. The interaction of structural and semantic biases in coherence and coreference. In *Poster at the 22nd Annual CUNY Conference on Sentence Processing*. Davis, CA.

Rohde, Hannah & Andrew Kehler. 2013. Grammatical and information-structural influences on pronoun production. Submitted for publication.

Rohde, Hannah & Andrew Kehler. in prep. Pronoun interpretation in a QUD model of discourse. In preparation.

Rohde, Hannah, Roger Levy & Andrew Kehler. 2011. Anticipating explanations in relative clause processing. *Cognition* 118. 339–358.

Smyth, R. J. 1994. Grammatical determinants of ambiguous pronoun resolution. *Journal of Psycholinguistic Research* 23. 197–229.

Stevenson, R., R. Crawley & D. Kleinman. 1994. Thematic roles, focusing and the representation of events. *Language and Cognitive Processes* 9. 519–548.

Stevenson, R., A. Knott, J. Oberlander & S. McDonald. 2000. Interpreting pronouns and connectives: Interactions among focusing, thematic roles, and coherence relations. *Language and Cognitive Processes* 15. 225–262.

Stevenson, Rosemary J., Alexander W. R. Nelson & Keith Stenning. 1995. The role of parallelism in strategies of pronoun comprehension. *Language and Speech* 38(4). 393–418.

Strawson, Peter F. 1964. Identifying reference and truth values. *Theoria* 30. 96–118.

Strube, Michael & Udo Hahn. 1999. Functional centering – grounding referential coherence in information structure. *Computational Linguistics* 25(3). 309–344.

Turan, Ümit Deniz. 1998. Ranking forward-looking centers in Turkish: Universal and language-specific properties. In Marilyn Walker, Aravind Joshi & Ellen Prince (eds.), *Centering in discourse*, 138–160. Oxford University Press.

Venditti, Jennifer J., Matthew Stone, Preetham Nanda & Paul Tepper. 2002. Discourse constraints on the interpretation of nuclear-accented pronouns. In *Proceedings of the 2002 international conference on speech prosody*. Aix-en-Provence, France.

Walker, Marilyn A., Masayo Iida & Sharon Cote. 1994. Japanese discourse and the process of centering. *Computational Linguistics* 20(2). 193–232.

Winograd, T. 1972. *Natural language understanding*. New York: Academic Press.